Metaproteomic evidence of changes in protein expression following a change in electrode potential in a robust biocathode microbiome

Dagmar H. Leary[1*], W. Judson Hervey, IV[1*], Anthony P. Malanoski[1], Zheng Wang[1], Brian J. Eddie[2], Gabrielle S. Tender[3], Gary J. Vora[1], Leonard M. Tender[1], Baochuan Lin[1], and Sarah M. Strycharz-Glaven[1]

[1]Center for Bio/Molecular Science and Engineering, Naval Research Laboratory, Washington, DC 20375, USA

[2]American Society for Engineering Education, Washington, DC 20036, USA

[3]California Institute of Technology, Pasadena, CA 91125, USA

*These authors contributed equally.

Received: 09-Dec-2014; Revised: 23-Jun-2015; Accepted: 05-Aug-2015.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/pmic.201400585.

ABSTRACT

Microorganisms that respire electrodes may be exploited for biotechnology applications if key pathways for extracellular electron transfer (EET) can be identified and manipulated through bioengineering. To determine whether expression of proposed Biocathode-MCL EET proteins are changed by modulating electrode potential without disrupting the relative distribution of microbial constituents, metaproteomic and 16S rRNA gene expression analyses were performed after switching from an optimal to suboptimal potential based on an expected decrease in electrode respiration. Five hundred and seventy-nine unique proteins were identified across both potentials, the majority of which were assigned to three previously defined Biocathode-MCL metagenomic clusters: a *Marinobacter* sp., a member of the family *Chromatiaceae*, and a *Labrenzia* sp. Statistical analysis of spectral counts using the Fisher's exact test identified 16 proteins associated with the optimal potential, five of which are predicted electron transfer proteins. The majority of proteins associated with the suboptimal potential were involved in protein turnover/synthesis, motility, and membrane transport. Unipept and 16S rRNA gene expression analyses indicated that the taxonomic profile of the microbiome did not change after 52 hours at the suboptimal potential. These findings show that protein expression is sensitive to the electrode potential without inducing shifts in community composition, a feature that may be exploited for engineering Biocathode-MCL.

INTRODUCTION

Meta-omics data provides a glimpse into the unique metabolic state of microorganisms while thriving from their associations with one another and with their natural surroundings. Microbial communities are able to respond to changing conditions they encounter in the environment, such as pH, nutrient concentrations, salinity or temperature.

Metaproteomics can be used to gain a functional understanding of this response [1], which is useful in identifying target genes for genetic manipulation to potentially engineer microbial communities for biotechnology applications. Microbiomes relevant to human health, energy and the environment, including the community in this study, are hard to precisely reproduce in the laboratory and can yield low quantities of biomass, creating difficulties for use of quantitative proteomic methods [2]. Despite these challenges, label-free techniques relying on statistical analysis of spectral counts have been demonstrated to provide accurate quantification of proteins from complex microbial communities [2, 3].

We recently reported metagenomic and metaproteomic analyses of an aerobic, non-phototrophic biocathode, referred to here as Biocathode-MCL after some of the most abundant and active cluster genomes noted: a *Marinobacter* sp., an unknown member of the family *Chromatiaceae*, and a *Labrenzia* sp. [4]. Biocathode-MCL was enriched from seawater over 5 years ago [5, 6] and uses a cathode electrode held at a set potential as an electron donor to fix CO_2 and generate biomass. A key outcome of previous analyses was that the *Chromatiaceae* appears to be a keystone organism, expressing proteins associated with carbon fixation and proteins hypothesized to be associated with EET [7], theoretically providing fixed carbon to non-autotrophic members of the community. Thus far, *Chromatiaceae* has not been cultivated outside of the electrode environment and heterotrophic isolates of *Marinobacter* and *Labrenzia* are unable to grow on the electrode without the addition of organic carbon [4]. An understanding of the EET and carbon flux associated with *Chromatiaceae*, *Marinobacter*, and *Labrenzia* is critical to engineer Biocathode-MCL to intentionally direct fixed carbon into secreted organic molecules, such as biofuels, that could be produced through microbial electrosynthesis [8-10].

In the present study we investigated the effect on protein expression of shifting the electrode potential more positive after Biocathode-MCL has fully formed. Previously

reported cyclic voltammetry (CV) of Biocathode-MCL [4, 6] indicated that current decreases when switching from 0.310 V vs. standard hydrogen electrode (SHE), an optimal potential used to grow Biocathode-MCL at the maximum rate of electrode respiration, to 0.470 V, a suboptimal potential at which the rate of electrode respiration is thought to be significantly reduced. Metaproteomic and 16S rRNA gene expression analyses were used to determine if changing the electrode potential induces 1) a change in expression of proteins associated with EET and CO_2 fixation predicted from the *Chromatiaceae* cluster genome, 2) a change in expression of potential EET proteins for other major biocathode constituents, and 3) a shift in the relative taxonomic distribution of the Biocathode-MCL microbiome due to the overall decrease in available energy for CO_2 fixation. Results show an electrochemical and physiological response by Biocathode-MCL after changing the electrode potential, as well as an overall robustness of the community indicated by no change to the relative abundance of microbial constituents over the duration of the experiment, features that may be exploited when designing engineered synthesis pathways.

MATERIALS AND METHODS

Biocathode biofilm cultivation.

An overview of the experimental workflow is depicted in Figure S1. All potentials reported here are versus standard hydrogen electrode (SHE). Two separate sets of four reactors (denoted as S1 for set 1 and S2 for set 2) were operated sequentially in order to demonstrate reproducibility of the electrochemical response, taxonomic distribution of the community, and protein expression from separately prepared inocula of Biocathode-MCL. Each set included two reactors for each potential from which samples were collected. For each set, the inoculum (ca. $2x10^5$ cells) was generated from a biocathode source electrode that is maintained at 0.310 V vs. SHE specifically for cultivation of the Biocathode-MCL

community (>5 years) [4, 6]. Metagenomes generated in previous experiments spanning several years have shown that the major Biocathode-MCL taxa and catalytic electrochemical features do not change [4, 6]. Therefore we consider new reactors inoculated from a source reactor to be biological replicates if electrochemical features and taxonomic distribution are similar to each other [4]. The number of cells in each inoculum was estimated using flow cytometry. A 3 cm x 3 cm section of carbon cloth from the source electrode was removed and the biofilm disrupted by vortexing for 30 sec, sonicating for 20 sec, and vortexing again for 30 sec in 10 mL of artificial seawater medium (ASW) medium. Cell extracts were stained with Syto9 nucleic acid stain (Life Technologies, Grand Island, NY) and a 50 μL volume was counted using an Accuri6 flow cytometer (BD Bioscience, San Jose, CA) set to slow fluidics, 488 nm excitation and 533/530 nm emission detection. Cells μL^{-1} were determined by taking the average number of events μL^{-1} of at least two counts compared to an ASW-carbon cloth control extract.

Bioelectrochemical reactors were identical to those previously described [4]. Reactors were 2L dual chambered microbial fuel cell reactors (Adams and Chittenden Scientific Glass) without membrane separation. Working electrodes were carbon cloth flags (length 6 cm, height 6 cm; total geometric surface area 36 cm^2 or 0.0036 m^2), and counter electrodes were graphite rods. Reactors were filled with sterile ASW, pH 6.5-6.8 [4, 6], and maintained at 30°C with stirring (VWR standard multi-position stirrer, setting "2" (150-200 rpm)) under atmospheric conditions. During biofilm growth, working electrodes were maintained at 0.310 V using a multichannel potentiostat (Solartron 1470E) under software control (Multistat, Scribner). After ca. 88 hours, CV was recorded (0.610 V to 0.260 V to 0.610 V, scan rate of 0.2 V/sec). Normalized CV is reported for each reactor by dividing the current of the cathodic scan by the peak current [11]. Following CV, the electrode potential was either returned to 0.310 V (Reactors 1 and 2, referred to as the "optimal potential") or increased to 0.470 V

(Reactors 3 and 4, referred to as the "suboptimal potential") for 52 hrs. Each biocathode was then removed from its reactor and sterile technique was used to divide the electrode into four sections: section 1 was used for RNA extraction for 16S rRNA gene analysis, section 2 was used for protein extraction and porcine trypsin digestion, section 3 was used for protein extraction and *Streptomyces erythraeus* trypsin (SET) digestion, and section 4 was used for cell counting by flow cytometry as described above.

Metaproteomics. The methods utilized for protein extraction and analysis were adapted from previous studies [4, 12-14], and are described below and an overview provided in Figure S2. Biocathode biofilm sample sections 2 and 3 from graphite cloth electrodes were submerged in B-PER Tris solution (ThermoFisher), sonicated on ice and centrifuged (4°C, 5,000 x *g*, 10 min) to sediment small graphite particles resulting from disrupting the electrode during sonication. Supernatants were transferred into new tubes and precipitated using 100 mM ammonium acetate in 100% MeOH. Precipitates were collected by centrifugation, dissolved in SDS-PAGE running buffer, separated by SDS-PAGE and distinct protein bands were cut, washed and destained. Bands from section 2 (n=7) were subjected to in-gel digestion overnight using sequencing grade modified porcine trypsin (PT) (Promega V511, Madison, WI). Bands from section 3 (n=9) were digested using *Streptomyces erythraeus* trypsin (SET, courtesy of Masaru Miyagi). Bands were cut in between visible protein bands, however the banding pattern was slightly different between PT and SET gels, thus the band number and size was adjusted to the banding pattern of the particular gel (gel images and cut patterns can be found deposited in the PRIDE partnership repository via ProteomeXchange with identifier PXD001590). Digested peptides were concentrated by centrifugal evaporation (SpeedVac, ThermoFisher) and reconstituted in 5% acetonitrile (ACN), 0.1% formic acid (FA) in H_2O (solvent A, Burdick & Jackson/Honeywell). Liquid chromatography tandem mass spectrometry (LC-MS/MS) was performed using a TempoMDLC system coupled to the

nanoelectrospray ionization source of a QSTAR Elite time-of-flight mass spectrometer (AB Sciex, Framingham, MA) as previously described [4, 12-14]. Samples were injected onto a reverse phase column (0.1 mm by 150 μm, 5 μm, 200Å Magic C18AQ particle size, Michrom, Auburn, CA) in 98% solvent A, 2% solvent B (95% ACN, 0.1% FA in H_2O). Separations were conducted at a flow rate of 700 nL/min for 30 min over a 60 min linear gradient of increasing solvent B by 0.5% per min to a final concentration of 60%

Tandem mass spectrometry data were extracted by AB Sciex MS Data Converter (v. 1.3 beta). Data were analyzed with both Mascot (Matrix Science, v. 2.4.1) and X!Tandem (thegpm.org; v. CYCLONE (2010.12.01.1)). All spectra acquired from bands in the same lane were merged prior to Mascot search. Mascot was set up to search the in-house biocathode metaproteomic database, previously created from Biocathode-MCL metagenomic sequencing [4], assuming the digestion enzyme trypsin. X!Tandem was set up to search a subset of the in-house built database also assuming trypsin. Mascot and X!Tandem were searched with parent and fragment mass tolerances of $\pm$0.30 Da. Mascot search variable modifications included deamidation of asparagine and glutamine, methionine oxidation, and carbamidomethyl cysteine. Variable modifications specified for X!Tandem search included amino terminal modifications (Glu->pyro-Glu, ammonia loss, and Gln->pyro-Glu), as well as asparagine and glutamine deamidation, methionine oxidation, and carbamidomethyl cysteine.

Scaffold (v. 4.2.1, Proteome Software, Portland, OR) applied delta mass correction and was used to validate tandem mass spectrometry data in the following manner. Identifications with peptide probability scores >95.0% and protein probability scores >90% (containing at least 1 peptide identification) by the Peptide Prophet [15] and Protein Prophet [16] algorithms, respectively, were considered detected for downstream analyses. Proteins containing similar peptides that could not be differentiated based on MS/MS analysis were

grouped to satisfy the principles of parsimony. When proteins sharing significant peptide evidence could not be discriminated they were grouped into clusters.

The Fisher's exact test (FET) (Scaffold v. 4.2.1) was used to statistically compare the weighted spectral counts generated by the Peptide Prophet algorithm between electrode potentials from each identified protein for each separate protein extraction, yielding a total of eight pooled samples for each electrode potential (Figure S2). For example, reactor 1 from set 1 (denoted S1R1) was extracted and analyzed twice, once using PT digestion and once using SET digestion. Use of weighted spectral counts ensures that each peptide is only counted once for the protein to which there is the strongest evidence of a match. Use of the FET has been demonstrated as an effective statistical approach to compare two conditions with small sample size when the dataset does not meet the assumptions of the t-test (normal distribution of data points, equal variance between conditions) [12]. A complete list of all spectral counts for each protein extract can be found in Table S3. The Benjamini-Hochberg (B-H) procedure [17] was applied to control for the false discovery rate (FDR) when multiple simultaneous comparisons are made (FDR set at 0.25) and results are denoted in Table 1 & 2. Additional statistical analysis of spectral counts using the beta binomial (BB) test [18] and the student's t-test (log-transformed spectral counts) is presented in Table S4 and S5 and results are denoted in Table 1 & 2.

All identified proteins were assigned to the genome clusters previously defined and described for Biocathode-MCL [4] and a complete list is available as supplemental material through the ProteomeXchange (http://www.ebi.ac.uk/pride/archive/) with identifier PXD001590 under file name "BiocathodAugandSep_allsamples_final". Protein identifications listed in Table 1 and Table 2 were abbreviated to the contig identifier and the gene number, e.g. Node_contig number_gene number. To infer protein function and metabolic pathways, amino acid sequences were annotated using the blastp algorithm and

NCBInr protein database with default settings (http://blast.ncbi.nlm.nih.gov). An annotation of the first highest scoring protein was accepted if the e-value <e-5. Annotations of proteins listed in Table 1 and Table 2 were further curated to check functional assignments for conserved domains and one error was found. Based on the top hit using blastp, Node_15807_1 was designated as a polymerase, but the conserved domain database (CDD) annotation lists this gene as encoding porin_2 (pfam02530). Functional assignments among hypothetical proteins were inferred from the CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml).

Unipept analysis. Independent analysis of identified peptide sequences was performed using Unipept 2.3 multipeptide analysis [19], leveraging the UniProt database (www.uniprot.org) and NCBI taxonomy. Briefly, the sequences of all identified peptides were submitted to the Unipept web application (http://unipept.ugent.be/) using the following settings to calculate the lowest common ancestors: multi-peptide analysis, peptides were deduplicated, isoleucine and leucine residues were equated, and advanced missed cleavage handling was applied.

RNA extraction. Biocathode biofilm RNA was extracted from section 1 as previously described [4]. Extracted RNA samples were subjected to Turbo DNase (Life Technologies) treatment according to the manufacturer's protocol and purified using the RNA Clean & Concentrator^{TM}-5 (Zymo Research) kit. The DNase treatment was repeated once per sample to ensure the removal of potential contaminating metagenomic DNA.

16S rRNA sequencing and phylogenetic analysis. Sixty to 160 ng total RNA per sample was reverse transcribed into cDNA using the PrimeScript RT-PCR kit (Takara, Mountain View, CA) according to the manufacturer's protocol. The V3 region of the 16S cDNA was amplified using the primers 357F (5'- ACTCCTACGGGAGGCAGCAG-3") and 518R (5'-ATTACCGCGGCTGCTGG-3') [20, 21]. Amplification reactions (25 μL) were assembled using 2.4 μL of cDNA as the PCR template. Amplification was carried out with an initial

denaturation at 95°C for 3 min; followed by 30 cycles of 95°C for 60 sec, 50°C for 60 sec, 72°C for 60 sec; followed by a final extension at 72°C, for 7 min. A secondary PCR amplification using index primers (Illumina, San Diego, CA) was performed and the amplicons were purified using the DNA Clean and Concentrator™-5 Kit (Zymo Research). Sequencing was conducted using a MiSeq™ instrument under automated software control (v. 2.2, Illumina. Raw FASTQ-formatted sequences were processed with the mothur software package (version 1.33.1) as detailed previously [22] with modifications described below. Paired-end reads from each sample were assembled into contiguous sequences and filtered using mothur [23]. Chimeric sequences were removed from the assemblies using the uchime program [24]. Non-ambiguous sequences ≥275 bp and ≥97th percentile were retained for analysis. The assembled reads were then processed with the RDP classifier to provide relative taxonomic abundances without assigning operational taxonomic units (OTUs) [25]. For comparison, OTUs were first assigned and classified using only mothur pipeline programs [23]. The relative taxonomic distribution of 16S rRNA reads or OTUs was rendered using the Krona tools software [26], which have been provided as supplemental html files along with metaproteomics data files.

Accession numbers. All metaproteomics data were deposited in the PRIDE partnership repository via ProteomeXchange with identifier PXD001590 (http://www.ebi.ac.uk/pride/archive/). Raw sequence reads were deposited to NCBI BioProject ID PRJNA244670, BioSample SAMN03166152. Raw FASTQ-formatted V3 16S rRNA sequencing data have been assigned NCBI BioSample accession SAMN03166152 accessible via BioProject PRJNA244670 [http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA244670].

RESULTS

Biocathode electrochemical features. Cyclic voltammetry previously indicated 0.310 V as the most positive potential at which Biocathode-MCL can sustain its maximum rate of electrode respiration [4, 6]. As the electrode potential becomes positive of 0.310 V to ca. 0.600 V, electrode respiration decreases until current becomes negligible [4, 6]. We therefore chose to investigate the effects of changing the electrode potential from 0.310 V (referred to here as the optimal potential) to 0.470 V (referred to here as the suboptimal potential), where electrode respiration is predicted to be ca. 50% maximum. Two sets (S1 and S2) of four identical reactors (R1, R2, R3, R4) were grown sequentially; reactors within the same set were inoculated at the same time from the same source electrode cell extract (Figure S1). All reactors were first grown at 0.310 V until current stabilized at ca. 88 hours (Figure 1) at which time CV was recorded (Figure S3). After CV, the electrode potential was switched to 0.470 V for R3 and R4, while R1 and R2 were maintained at 0.310 V for an additional 52 hours (Figure 1).

In Figure 1 current indicates the instantaneous rate of EET by Biocathode-MCL, where electrons are presumably used to reduce the terminal electron acceptor (i.e., O_2), fix CO_2, or accumulate in reduced electron transfer proteins or other cofactors (e.g. quinones) in cells or the biofilm matrix. Despite efforts to minimize biological variability, the maximum current achieved by each reactor varied widely, highlighting the inherent difficulty in reproducing complex microbiomes. Maximum current peaks in chronoamperometry (CA) (Figure 1; Table S1) are suspected to be indicative of growth limitation due to equilibrium of O_2 in the biofilm or other chemical gradient limitations (e.g. H+). Hysteresis in CV during the anodic scan (Figure S3) also shows a feature that may be associated with a diffusion limited electron transfer process, such as the diffusion of O_2 back into the biofilm following the cathodic scan. Nonetheless, the shape of the normalized cathodic current of all reactors is

consistent (Figure S4), taken here to indicate that the same mechanism of electrode respiration occurs in each replicate.

All reactors experienced an immediate spike in current lasting ca. 20 minutes when CA was resumed following CV (Figure 1). Reactors adjusted to 0.470 V experienced an additional current peak analogous to that observed after inoculation between 3 and 26 hours following CV, indicating some change in electrode respiration had occurred due to the change in electrode potential. No additional current peak was observed in reactors returned to 0.310 V following CV other than the initial spike, indicating that CV alone did not affect the additional current peak in suboptimal potential reactors.

Metaproteomes from optimal and suboptimal electrode potentials. Previous protein extraction yields from Biocathode-MCL were low [4], exemplifying a frequently overlooked aspect of sample preparation: processing limited quantities of protein starting material [27]. Therefore we chose to digest protein samples with two different types of trypsin: PT and SET. As described previously [28], SET has greater amidase activity than modified PT and it lacks autolytic activity, thus reducing sample contamination with autotryptic peptides. Such properties are advantageous for use in preparing low yield samples, as well as complex environmental samples for which there may be limited metagenomic sequencing information for protein identification. To this end, we leveraged SET's advantages and demonstrated its first application on a complex metaproteome. The use of SET combined with an increased number of cut gel bands for the second set of protein extracts from each reactor led to an additional 310 protein identifications and increased spectrum assignments by up to 28% (Table S2).

Protein identifications and cluster genome assignments were made using the previously described draft metagenome of Biocathode-MCL [4] resulting in a total of 579 unique proteins across all eight biocathode protein extractions. Four hundred and three of

these proteins were identified in electrode samples from both potentials. One hundred and one of these proteins were identified from electrode samples at the optimal potential only and 75 proteins were identified from the suboptimal potential only. Proteins identified as 'optimal only' were manually curated to remove those predicted to be part of a protein complex if other subunits from the same complex (predicted from the same metagenomic contig) were observed at the suboptimal potential, and vice versa for those identified as 'suboptimal only' [4]. Additionally, 'suboptimal only' proteins were manually curated to remove proteins that were previously identified at the optimal potential [4]. The overall distribution of total proteins identified from all reactors among proposed cluster genomes was similar to that previously reported for Biocathode-MCL [4], where *Marinobacter*, *Chromatiaceae*, and *Labrenzia* were the most highly represented (Figure 2). Although *Marinobacter* and *Labrenzia* cluster genomes had roughly the same number of total identified proteins, *Labrenzia* had a greater number of proteins that were uniquely identified at either the optimal or suboptimal potential.

Due to a combination of low sample biomass, limited protein extraction yield, and the considerable expenditure of resources required for biological replication (e.g. increasing the number of reactors), a label-free method was applied to estimate the difference in abundance of proteins from samples collected at the optimal and suboptimal electrode potentials. The Fisher's exact test (FET) was used to statistically compare spectral counts across all protein extractions from each condition (Table 1 & 2). SET and PT extracts from the same sample were analyzed separately in order to highlight the difference in protein identification between the two digestions. In order to further substantiate differences in protein expression identified by the FET, additional statistical tests were carried out. The Benjamini-Hochberg (B-H) procedure was applied to the FET p-values, and the beta binomial (BB) test [18] and student's t-test (log-transformed spectral counts) were also carried out considering PT and

SET digestions separately for consistency with the FET (Tables S4 and S5). The BB test is a less commonly used non-parametric test that accounts for both intra- and inter-sample variability. The student's t-test is a common statistical test for samples with normal data distribution and equal variance, but can be applied for spectral counting if data are transformed [18]. Proteins identified as significantly associated with the optimal or suboptimal potential using the BB test and student's t-test in addition to the FET are denoted in Table 1 & 2, and a complete list of results (including spectral counts) can be found in Tables S4 and S5. The results of the FET can be summarized as follows:

1. Five enzymes were predicted to be associated with the optimal electrode potential. Only one, a predicted quinoprotein alcohol dehydrogenase from *Marinobacter*, was significantly different by the BB test and t-test.
2. Eight structural and/or hypothetical proteins and several ABC transporters from *Alcanivoraceae*, *Chromatiaceae*, *Labrenzia*, *Marinobacter* and *Kordiimonas* were associated with the optimal potential. A hypothetical protein from *Kordiimonas* was also significant by the BB test and t-test.
3. Two ABC transporters from *Labrenzia* and one from *Marinobacter* were significantly associated with the optimal potential. One of these, a sugar ABC transporter from *Labrenzia*, was also significant by the BB test and t-test.
4. Five proteins from *Marinobacter* and *Chromatiaceae* involved in protein expression, turnover, and interaction were associated with the suboptimal samples.
5. Two proteins associated with motility/trafficking were identified from *Marinobacter* at the suboptimal potential. One was a flagellin protein and was also significant by the BB test and t-test.
6. A number of transporter and receptor proteins from *Chromatiaceae*, *Labrenzia*, *Marinobacter*, and *Parvibaculum* were associated with the suboptimal potential. A

Marinobacter membrane protein with a receptor domain for Fe^{3+} was also significant by the BB test and t-test.

7. A number of hypothetical proteins were identified from *Marinobacter, Chromatiaceae, Parvibaculum, and Rhodospirillaceae* cluster genomes at the suboptimal potential. One with a predicted flagella domain from *Parvibaculum* was also significant by the BB test and t-test.

In general, a greater number of redox active proteins were associated with the optimal electrode samples, while a greater number of transporter and structural proteins, including flagella-related proteins, were associated with the suboptimal samples. At this time we cannot further infer the role of these proteins in either EET or carbon flux in Biocathode-MCL but note that results suggest a change in protein expression occurs in at least some constituents when the electrode is changed to 0.470 V.

Unipept analysis and 16S rRNA gene expression.

In order to determine whether changing the electrode potential induced a shift in the overall taxonomic distribution of the biocathode biofilm, phylogenetic analysis of matched peptide sequences was performed using the Unipept software package (Figure 3). Unipept was able to assign ≥90% of matched peptide sequences to *Bacteria* for all electrode protein samples from both the optimal and suboptimal reactors (Table S6). Forty percent or greater of matched peptides pooled across all reactors from a given potential were assigned to *Gammaproteobacteria,* with the majority identified as belonging to the order *Alteromonadales* (contains genus *Marinobacter*) (Table S6). Twenty-nine percent or greater of matched peptides pooled across the same reactors were assigned *Alphaproteobacteria* and consisted almost entirely of peptides matching *Rhodobacterales* (contains genus *Labrenzia*) (Table S6). Relative abundance of *Alpha-* and *Gammaproteobacteria,* predicted by peptide taxonomic assignment varied between individual reactors, even those in the same set (Table

S7 and Figure S5 & S6). However, when averaged across samples from reactors at the same potential, there was no significant difference in the percentage of peptides assigned to these groups (Table S7). Contrary to previous results based on metagenomic data [4] as well as the distribution of predicted proteins in Figure 2, members of the order *Chromatiales* (order to which *Chromatiaceae* belongs) represented <1% of peptide assignments at either electrode potential.

In addition to Unipept analysis, the relative taxonomic distribution of the Biocathode-MCL microbiome at the optimal and suboptimal potentials was estimated by transcript abundance of the 16S rRNA gene V3 region between the optimal and suboptimal electrodes (Figure S7 & S8). When reads were grouped using the RDP classifier (Figure S7), *Gammaproteobacteria*, specifically the families *Alteromonadaceae* contains genus *Marinobacter*) and *Ectothiorhodospiraceae* (close relative of *Chromatiaceae*) made up 71% of reads at the optimal potential and 67% of reads at the suboptimal potential (Table S7). The remaining reads were mostly identified as *Alphaproteobacteria*, with the majority from the family *Rhodobacteriaceae* (contains genus *Labrenzia*). Read classification with the mothur workflow (Figure S8) was consistent in terms of the percent matching *Gammaproteobacteria* and *Alphaproteobacteria*, however, *Chromatiales* were not identified and presumed to make up the majority of unclassified *Gammaproteobacteria*. As with Unipept analysis, the average distribution of taxa based on 16S rRNA gene V3 region sequencing was not statistically different between the optimal and suboptimal potential (Table S7). Variability in identified taxa between each reactor was similar with that observed between reactors using Unipept (Table S7; supplemental plots for 16S rRNA gene expression analysis from all reactors and both classification methods can be found in the PRIDE repository).

DISCUSSION

Overall, the distribution of total proteins identified from both the 0.310 V and 0.470 V electrode potentials among defined cluster genomes was similar to that previously reported for Biocathode-MCL at 0.310 V [4]. Unipept and 16S rRNA gene expression analyses showed that the relative abundance of biocathode constituents did not change following a change in electrode potential. Taxonomic identifications made using Unipept, RDP and mothur are generally in agreement at the class level, however, Unipept and mothur were unable to assign peptides or 16S rRNA gene transcripts belonging to *Chromatiales* or *Chromatiaceae*. Protein identifications based cluster genomes indicate *Chromatiaceae* is highly active and are consistent with previous estimates determined by metaproteomic analysis and predicted relative abundance of this taxa based on RDP classifications of the 16S rRNA gene transcripts [4]. All taxonomic assignment methods used rely on curated reference databases that provide accurate resolution when a closely related species is present in the database and uncharacterized organisms, such as those from environmental enrichments, may be not be classified or may be misidentified. From our previous work, we know that the *Chromatiaceae* cluster genome represents a new organism not closely related to any organisms available in the reference databases used here.

Electrochemical analysis used to characterize the rate of EET before and after changing the electrode potential indicated: 1) a wide variability in sustained current among otherwise identical biological replicates, 2) a second current maximum observed after changing the potential to 0.470 V that was not observed when the potential was maintained at 0.310 V, indicating a change in electrode respiration due to the change in electrode potential; and 3) current not trending to ~50% after changing the electrode potential. The first result may reflect variations in the relative abundance and orientation of biocathode constituents at the electrode surface between reactors due to unknown factors affecting spatial and temporal electrode colonization. The number of extractable cells at the end of the experiment did not

correlate with current (Table S1), and therefore is not expected to be the reason for this discrepancy.

The second and third results are surprising because they indicate that an increase in electrode potential resulted in an increase rather than the expected decrease in rate of electrode respiration. The expected outcome was based on CV however, which may have been too short of a time frame to assess long terms effects of maintaining a set potential. One possible explanation for the observed second current maximum may be a temporary increase in the rate of EET required to obtain a sufficient amount of energy per unit time at the lower driving force of the higher potential electrode. We previously proposed that an outer membrane protein, likely a *c*-type cytochrome, could be the terminal oxidase, feeding into the electron transport chain to drive NAD+ reduction to NADH. We identified a homolog of Cyc2 from *Acidithiobacillus ferrooxidans* [29], which has a predicted formal potential of 0.560 V at pH 4.8 [30]. Even if we consider a decrease in formal potential at circumneutral pH of 0.1-0.2 V, EET through a Cyc2-like protein could still feasibly carry out EET at 0.470 V. The amount of energy obtained by reducing O_2 will be lower for each electron acquired from the electrode, which may reduce the efficiency of proton pumping across the membrane by the *cbb*$_3$ oxidase. Furthermore, it would also require a greater proton motive force to drive electrons uphill (energy consuming) to regenerate NADH for CO_2 fixation. The ratio of electrons that can go downhill (energy generating) to those that go uphill would need to be approximately twice as much at the higher electrode potential than at the lower electrode potential [31]. Proteomics data gave no indication of any new proteins or an increase in the abundance of existing electron transport proteins at the suboptimal potential and the increase in current was not sustainable. One explanation may be that inner membrane proteins needed for respiratory flexibility at varying potentials are constitutively expressed, as has recently been shown for *Geobacter sulfurreducens* [32], and therefore are

not apparent in protein comparison between the two conditions. Down-regulation of the CBB cycle that would be associated with electrons being redirected to respiration was also not apparent from proteomics analysis as proteins essential for carbon fixation were identified in both the optimal and suboptimal reactors.

The FET has been shown to effectively detect differentially expressed proteins based on spectral counts between two conditions given a low number of samples, non-normally distributed data, and unequal variance between samples [33]. When the FET was applied to compare spectral counts from the optimal and suboptimal potential samples, a greater number of redox proteins were observed at the optimal potential and a greater number of structural and receptor/transporter proteins were observed at the suboptimal potential. P-value adjustment using the B-H procedure predicts that only 4 proteins associated with the optimal potential and only 3 associated with the suboptimal potential can confidently be considered significant using the FET. Although experimental validation is required, a conservative interpretation of these results, in combination with the noted electrochemical response, is that changing the electrode potential 0.160 V positive initiates a direct metabolic response in at least some biocathode constituents as indicated by association of proteins for turnover/synthesis, motility, and transport of small molecules with the suboptimal potential. Reactors switched to 0.470 V never achieved a steady state current prior to sampling, limiting our ability to draw conclusions regarding proteins important for respiration at this potential.

Only three proteins at the optimal potential and three proteins at the suboptimal potential indicated as significantly different by the FET were also found to be significantly different using the BB test and the t-test. Three of these 6 belong to the *Marinobacter* cluster genome and may indicate that this biocathode constituent is sensitive to the changing electrode potential. Two of these proteins were associated with the suboptimal potential, a flagellin protein and a membrane protein with a Fe^{3+} receptor domain, and may be involved

in sensing changes to the extracellular environment. The one protein associated with the optimal potential, a quinoprotein alcohol dehydrogenase, is known to be involved in membrane electron transfer [34], indicating a possible target for further analysis of ET in *Marinobacter*.

CONCLUSIONS

Metaproteomics combined with label-free quantification were successfully used here to determine if proteins identified from Biocathode-MCL were different when the electrode potential was changed. Biocathode-MCL was able to tolerate a 0.160 V shift positive in electrode potential over two days with no change to community composition despite electrochemical and proteomic evidence of a biological response. The overall stability of the community following perturbation of the driving force for respiration and carbon fixation implies a robust energy metabolism for Biocathode-MCL. Ongoing work includes further characterization of proteins identified from *Marinobacter* to be associated with changing the electrode potential. Identification of functional proteins from Biocathode-MCL whose expression may be modulated by the electrode potential will be used to engineer the biocathode microbial community for biotechnology applications.

ACKNOWLEDGEMENTS

We thank the DoD High Performance Computing Modernization Program's (HPCMP) PETTT staff at the Naval Research Laboratory for assistance with software configuration. We thank Dr. Masaru Miyagi from Case Western Reserve University, Cleveland, OH for kindly sharing purified SET. We thank Dr. Robert W. Li, USDA, for initial sequencing and annotation of the Biocathode-MCL metagenome. We thank the PRIDE

Team for assistance with processing submission of mass spectrometry data to the ProteomeXchange Consortium.

This work was funded by the Office of Naval Research via U.S. NRL core funds, as well as under the following award numbers to S.S.-G.: N0001413WX20995, N0001414WX20485, N0001414WX20518, and N0001415WX00195. The opinions and assertions contained herein are those of the authors and are not to be construed as those of the U.S. Navy, military service at large, or U.S. Government.

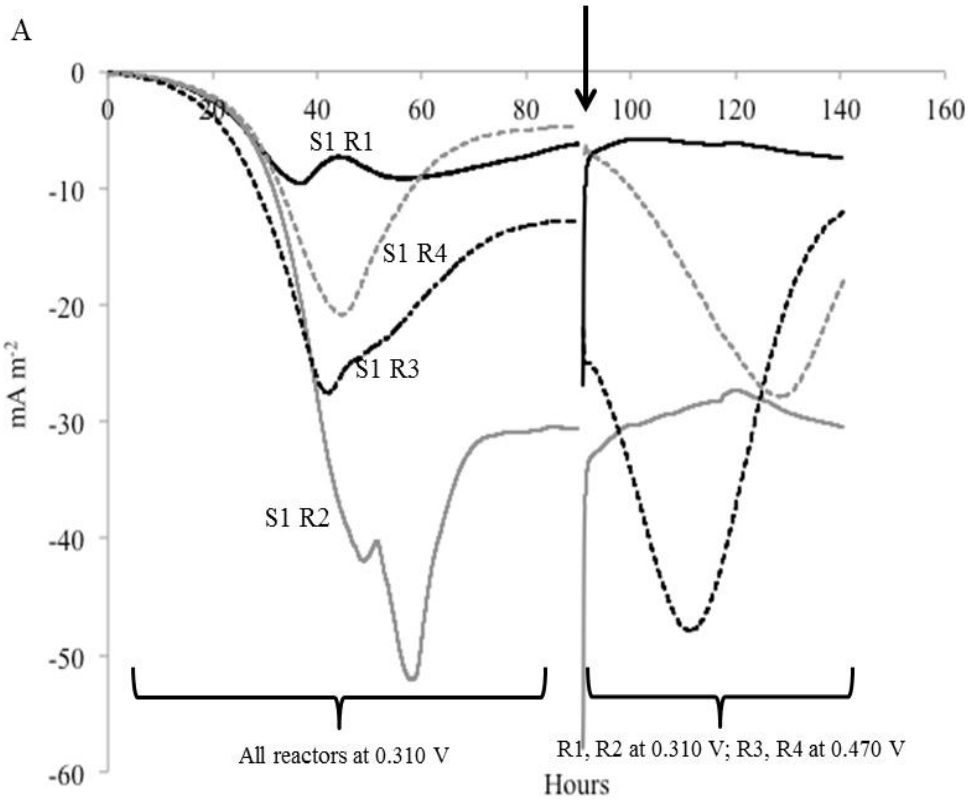
REFERENCES

[1] Banfield, J. F., Verberkmoes, N. C., Hettich, R. L., Thelen, M. P., Proteogenomic approaches for the molecular characterization of natural microbial communities. *OMICS* 2005, *9*, 301-333.
[2] VerBerkmoes, N. C., Denef, V. J., Hettich, R. L., Banfield, J. F., SYSTEMS BIOLOGY Functional analysis of natural microbial consortia using community proteomics. *Nature Reviews Microbiology* 2009, *7*, 196-205.
[3] von Bergen, M., Jehmlich, N., Taubert, M., Vogt, C., *et al.*, Insights from quantitative metaproteomics and protein-stable isotope probing into microbial ecology. *ISME J.* 2013, *7*, 1877-1885.
[4] Wang, Z. W., Leary, D. H., Malanoski, A. P., Li, R. W., *et al.*, A previously uncharacterized, non-photosynthetic member of the Chromatiaceae is the primary CO2 fixing constituent in a self-regenerating biocathode. *Appl. Environ. Microbiol.* 2015.
[5] Malik, S., Drott, E., Grisdela, P., Lee, J., *et al.*, A self-assembling self-repairing microbial photoelectrochemical solar cell. *Energy Environ. Sci.* 2009, *2*, 292-298.
[6] Strycharz-Glaven, S. M., Glaven, R. H., Wang, Z., Zhou, J., *et al.*, Electrochemical Investigation of a Microbial Solar Cell Reveals a Nonphotosynthetic Biocathode Catalyst. *Appl. Environ. Microbiol.* 2013, *79*, 3933-3942.
[7] Singer, E., Heidelberg, J. F., Dhillon, A., Edwards, K. J., Metagenomic insights into the dominant Fe(II) oxidizing Zetaproteobacteria from an iron mat at Lo ihi, Hawai l. *Front. Microbiol.* 2013, *4*, 52.
[8] Rabaey, K., Rozendal, R. A., Microbial electrosynthesis - revisiting the electrical route for microbial production. *Nat. Rev. Microbiol.* 2010, *8*, 706-716.
[9] Lovley, D. R., Nevin, K. P., Electrobiocommodities: powering microbial production of fuels and commodity chemicals from carbon dioxide with electricity. *Curr. Opin. Biotechnol.* 2013, *24*, 385-390.
[10] Marshall, C. W., Ross, D. E., Fichot, E. B., Norman, R. S., May, H. D., Electrosynthesis of commodity chemicals by an autotrophic microbial community. *Appl. Environ. Microbiol.* 2012, *78*, 8412-8420.

[11] Strycharz-Glaven, S. M., Tender, L. M., Study of the Mechanism of Catalytic Activity of G. Sulfurreducens Biofilm Anodes during Biofilm Growth. *Chemsuschem* 2012, *5*, 1106-1118.
[12] Leary, D. H., Hervey, W. J. t., Deschamps, J. R., Kusterbeck, A. W., Vora, G. J., Which metaproteome? The impact of protein extraction bias on metaproteomic analyses. *Mol. Cell. Probes* 2013, *27*, 193-199.
[13] Leary, D. H., Hervey, W. J. t., Li, R. W., Deschamps, J. R., *et al.*, Method development for metaproteomic analyses of marine biofilms. *Anal. Chem.* 2012, *84*, 4006-4013.
[14] Leary, D. H., Li, R. W., Hamdan, L. J., Hervey, W. J. t., *et al.*, Integrated metagenomic and metaproteomic analyses of marine biofilm communities. *Biofouling* 2014.
[15] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 2002, *74*, 5383-5392.
[16] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003, *75*, 4646-4658.
[17] Benjamini, Y., Hochberg, Y., Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 1995, *57*, 289-300.
[18] Pham, T. V., Piersma, S. R., Warmoes, M., Jimenez, C. R., On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics* 2010, *26*, 363-369.
[19] Mesuere, B., Devreese, B., Debyser, G., Aerts, M., *et al.*, Unipept: Tryptic Peptide-Based Biodiversity Analysis of Metaproteome Samples. *J. Proteome Res.* 2012, *11*, 5773-5780.
[20] Humblot, C., Guyot, J. P., Pyrosequencing of Tagged 16S rRNA Gene Amplicons for Rapid Deciphering of the Microbiomes of Fermented Foods Such as Pearl Millet Slurries. *Appl. Environ. Microbiol.* 2009, *75*, 4354-4361.
[21] Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G., Neufeld, J. D., Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads (vol 77, pg 3846, 2011). *Appl. Environ. Microbiol.* 2011, *77*, 5569-5569.
[22] Gregoire, K. P., Glaven, S. M., Hervey, W. J. t., Lin, B., Tender, L. M., Enrichment of a high-current density denitrifying microbial biocathode. *J. Electrochem. Soc.* 2014, *161*, H3049-H3057.
[23] Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., *et al.*, Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* 2009, *75*, 7537-7541.
[24] Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., Knight, R., UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011, *27*, 2194-2200.
[25] Wang, Q., Garrity, G. M., Tiedje, J. M., Cole, J. R., Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 2007, *73*, 5261-5267.
[26] Ondov, B. D., Bergman, N. H., Phillippy, A. M., Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 2011, *12*.
[27] Hervey, W. J., Strader, M. B., Hurst, G. B., Comparison of digestion protocols for microgram quantities of enriched protein samples. *J. Proteome Res.* 2007, *6*, 3054-3061.
[28] Kiser, J. Z., Post, M., Wang, B. L., Miyagi, M., Streptomyces erythraeus Trypsin for Proteomics Applications. *J. Proteome Res.* 2009, *8*, 1810-1817.
[29] Yarzabal, A., Brasseur, G., Ratouchniak, J., Lund, K., *et al.*, The high-molecular-weight cytochrome c Cyc2 of Acidithiobacillus ferrooxidans is an outer membrane protein. *J. Bacteriol.* 2002, *184*, 313-317.

[30] Castelle, C., Guiral, M., Malarte, G., Ledgham, F., *et al.*, A new iron-oxidizing/O-2-reducing supercomplex spanning both inner and outer membranes, isolated from the extreme acidophile Acidithiobacillus ferrooxidans. *J. Biol. Chem.* 2008, *283*, 25803-25811.
[31] Roger, M., Castelle, C., Guiral, M., Infossi, P., *et al.*, Mineral respiration under extreme acidic conditions: from a supramolecular organization to a molecular adaptation in Acidithiobacillus ferrooxidans. *Biochem. Soc. Trans.* 2012, *40*, 1324-1329.
[32] Levar, C. E., Chan, C. H., Mehta-Kolte, M. G., Bond, D. R., An Inner Membrane Cytochrome Required Only for Reduction of High Redox Potential Extracellular Electron Acceptors. *mBio* 2014, *5*, e02034-02014.
[33] Zhang, B., VerBerkmoes, N. C., Langston, M. A., Uberbacher, E., *et al.*, Detecting differential and correlated protein expression in label-free shotgun proteomics. *J. Proteome Res.* 2006, *5*, 2909-2918.
[34] Ikeda, T., Kano, K., Bioelectrocatalysis-based application of quinoproteins and quinoprotein-containing bacterial cells in biosensors and biofuel cells. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2003, *1647*, 121-126.
[35] Wargo, M. J., Homeostasis and Catabolism of Choline and Glycine Betaine: Lessons from Pseudomonas aeruginosa. *Appl. Environ. Microbiol.* 2013, *79*, 2112-2120.
[36] Galinski, E. A., Truper, H. G., Betaine, a Compatible Solute in the Extremely Halophilic Phototropic Bacterium Ectothiorhodospira-Halochloris. *FEMS Microbiol. Lett.* 1982, *13*, 357-360.

Figure 1. Chronoamperometry (CA) from A) Set 1 (S1) and B) Set 2 (S2) reactors: R1 (solid black line), R2 (solid gray line), R3 (dashed black line), and R4 (dashed gray line). Arrow indicates when CV (depicted in Figure S3) was recorded for all reactors, and the electrode potential of R3/R4 reactors was adjusted to 0.470 V.



B

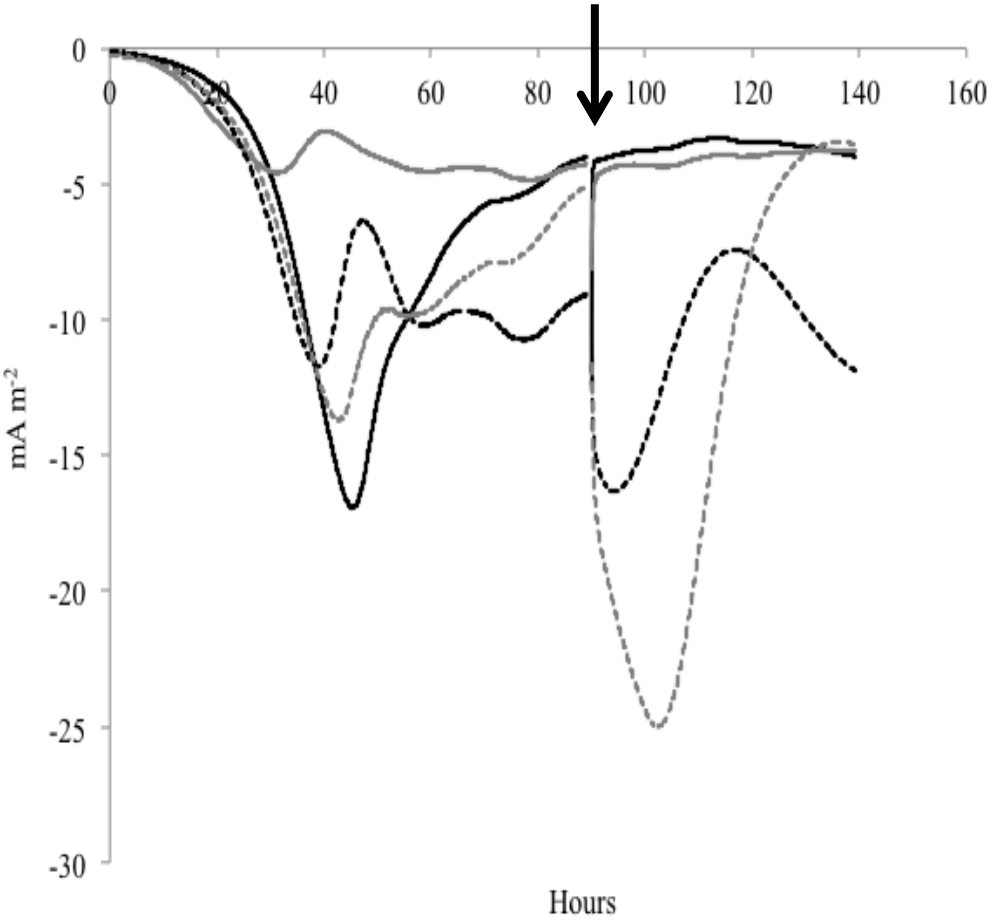


Figure 2. Number of proteins identified in each genome cluster predicted from previous metagenomic analysis of Biocathode-MCL.

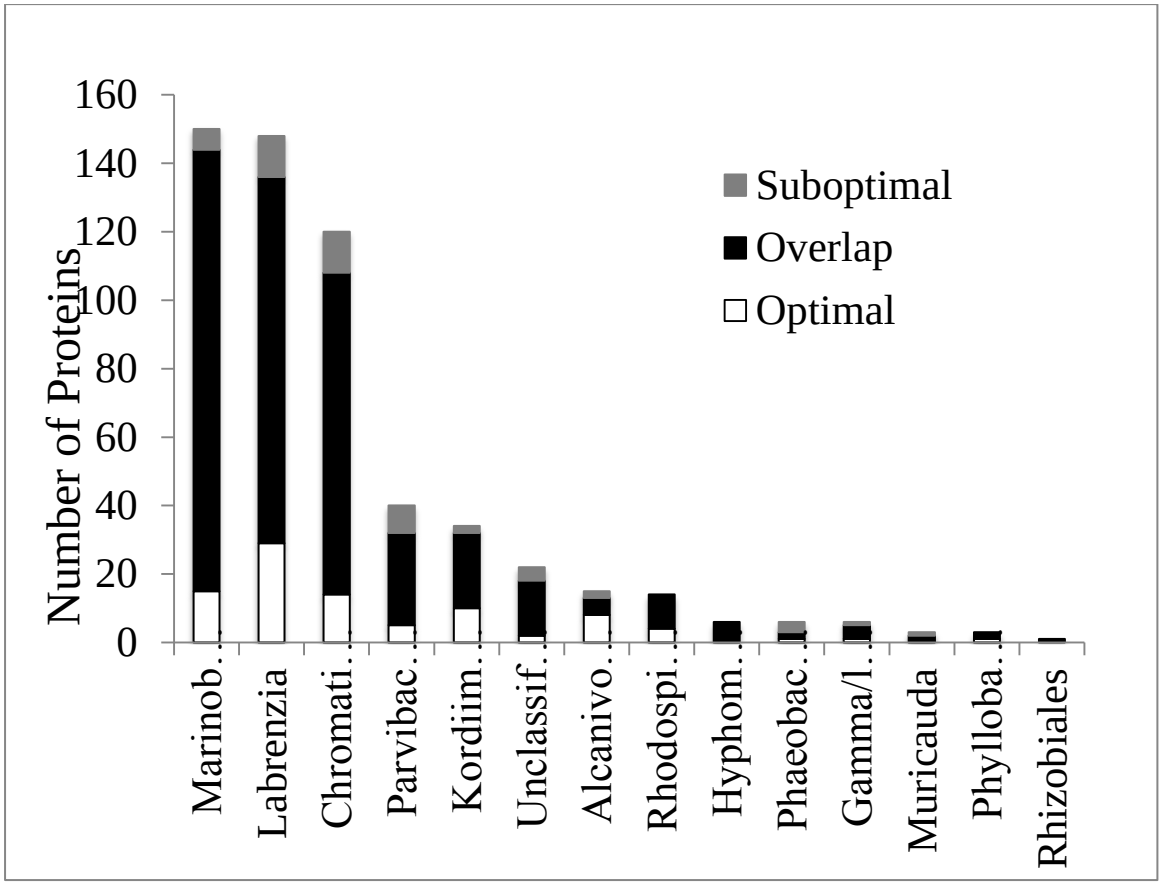


Figure 3. Relative abundance of peptide phylogenetic assignment by Unipept analysis. A) Optimal: 670 out of 1063 peptides were matched (peptides were deduplicated, I and L residues were equated, advanced missed cleavage handling). B) Suboptimal: 670 out of 1063 peptides were matched (peptides were deduplicated, I and L residues were equated, advanced missed cleavage handling).

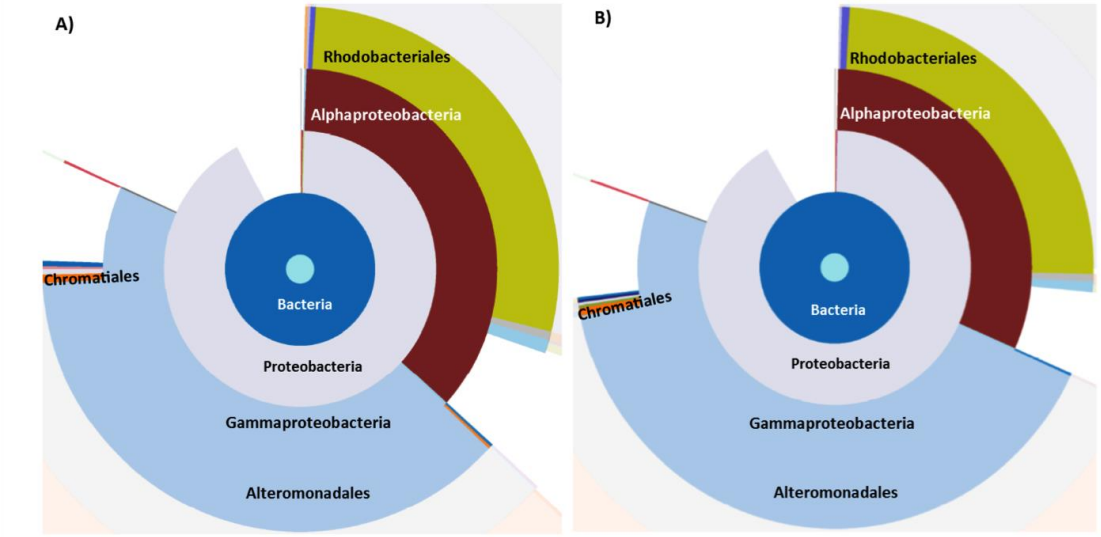


Table 1. Proteins identified from biocathode samples with statistically greater number of spectral counts when operated at the optimal potential (0.310 V) only.

ORF ID	NCBI ANNOTATION	Cluster genome	Fisher p-value	Functional category
Enzymes*				
NODE_837_61	Cytochrome C class I	*Chromatiaceae*	0.0042	electron transfer
NODE_277_255	electron transfer flavoprotein subunit beta	*Kordiimonas*	0.0089	electron transfer
NODE_728_49	electron transfer flavoprotein subunit beta	*Labrenzia*	0.028	electron transfer
NODE_2140_9	quinoprotein alcohol dehydrogenase	*Marinobacter*	**<0.0001**^	
NODE_2170_44	methionine sulfoxide reductase	*Marinobacter*	0.028	oxidative stress regulation
Structural and Hypothetical				
NODE_16387_1	flagellin	*Alcanivoraceae*	0.028	motility/trafficking
NODE_2320_41	peptidoglycan-binding protein	*Chromatiaceae*	**<0.0001**	sugar binding protein
NODE_837_51	hypothetical	*Chromatiaceae*	0.011	unknown
NODE_277_323	hypothetical	*Kordiimonas*	0.0066^	unknown
NODE_181_44	hypothetical, flagellin domain	*Kordiimonas*	0.026	motility/trafficking
NODE_3683_49	peptide binding protein	*Marinobacter*	0.0063	
NODE_240_101	50S ribosomal protein L9	*Labrenzia*	0.047	protein synthesis
NODE_15807_1	porin	*Labrenzia*	0.047	
Transporters and Receptors				

NODE_22_72	sugar ABC transporter	*Labrenzia*	**<u><0.0001^</u>**	ABC transporter
NODE_893_1	ABC transporter substrate-binding protein	*Labrenzia*	0.028	ABC transporter
NODE_1148_52	C4-dicarboxylate ABC transporter	*Marinobacter*	**0.00067**	ABC transporter

*Includes electron transfer proteins.

Fisher p-values in **bold** indicate significant by Benjamini-Hochberg procedure (FDR 25%). <u>Underlined</u> values were significant (p-value <0.05) using the beta-binomial test (BB test). ^indicates significant (p-value <0.05) by t-test on log transformed spectral counts.

Table 2. Proteins identified from biocathode samples with statistically greater number of spectral counts after operating at the suboptimal electrode potential (0.470 V) for 52 hours.

ORF ID	NCBI ANNOTATION	Cluster genome	Fisher p-value	Functional category
Enzymes				
NODE_5518_1	glycine betaine transmethylase	*Parvibaculum*	0.0086	
Structural and Hypothetical				
NODE_1547_6	50S ribosomal protein L7/L12	*Chromatiaceae*	0.0081	ribosome, protein synthesis
NODE_2048_7	30S ribosomal protein S1	*Chromatiaceae*	0.023	ribosome, protein synthesis
NODE_2943_37	molecular chaperone DnaK	*Chromatiaceae*	**0.0026**	chaperone, protein turnover
NODE_307_76	membrane protein	*Chromatiaceae*	0.023	unknown
NODE_83_108	hypothetical protein	*Chromatiaceae*	0.023	protein-protein interaction domain
NODE_1775_17	lipoprotein	*Marinobacter*	0.023	
NODE_403_5	flagellin	*Marinobacter*	**<0.0001**	motility/trafficking
NODE_476_49	50S ribosomal protein L15	*Marinobacter*	0.035	ribosome, protein synthesis
NODE_508_70	hypothetical protein HP15_1830	*Marinobacter*	0.028	unknown
NODE_6203_4	fimbrial protein	*Marinobacter*	0.04	motility/trafficking
NODE_2368_40	hypothetical protein Plav_2552, flagella domain	*Parvibaculum*	0.0052^	unknown
NODE_1848_121	hypothetical protein, partial	*Rhodospirillaceae*	0.0037	unknown
Transporters and Receptors				
NODE_6881_4	ligand-gated channel protein	*Chromatiaceae*	0.025	
NODE_1173_134	spermidine/putrescine ABC transporter substrate-binding protein	*Labrenzia*	0.0072	ABC transporter
NODE_3258_14	peptide ABC transporter substrate-binding protein	*Labrenzia*	0.014	ABC transporter
NODE_1573_14	membrane protein, domain for outer membrane receptor for Fe3+-dicitrate	*Marinobacter*	**0.0025^**	
NODE_2210_36	branched-chain amino acid ABC transporter	*Marinobacter*	0.033	ABC transporter, amino acid transport

NODE_3683_20	amino acid ABC transporter substrate-binding domain	*Marinobacter*	0.028	ABC transporter, amino acid transport
NODE_522_69	membrane protein	*Marinobacter*	0.042	unknown
NODE_2533_8	TonB-dependent receptor plug	*Parvibaculum*	0.032	TonB receptor

Fisher p-values in **bold** indicate significant by Benjamini-Hochberg procedure (FDR 25%). Underlined values were significant (p-value <0.05) using the beta-binomial test (BB test). ^indicates significant (p-value <0.05) by t-test on log transformed spectral counts.